

A novel method for estimating propagation pathloss in millimeter-wave communication systems

Vu Thanh Quang¹, Do Huu Duc¹, Vu Van Yem¹, Hoang Thi Phuong Thao²

¹Department of Communication Engineering, School of Electrical and Electronic Engineering, Hanoi University of Science and Technology, Hanoi, Vietnam

²Department of Electronics and Telecommunications, Electric Power University, Hanoi, Vietnam

Article Info

Article history:

Received Apr 30, 2023

Revised May 15, 2023

Accepted Jul 13, 2023

Keywords:

Clustering technique

Co-IEE algorithm

Millimeter-wave communication

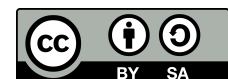
Minimum mean square error method

Pathloss estimation

ABSTRACT

In millimeter wave communication systems, the construction of a path loss model that is close to empirical one plays an important role in planning coverage, system capacity, and link budgets. In this paper, we propose a novel approach applied unsupervised machine learning for propagation pathloss estimation in millimeter wave communication systems based on the iterative procedure of cooperative and iterative evaluation exchange (Co-IEE) algorithm. The propagation path loss models of the time city—an urban area in the center of Hanoi, Vietnam in both line-of-sight (LOS) and non line-of-sight (NLOS) are estimated and calculated using the proposed approach and then they are compared to the minimum mean square error (MMSE). The estimated results of the proposed model show the approximation with the optimum one returned by MMSE method. Moreover, the proposed model of estimating path loss can solve the problem of sensitivity with outliers existing in MMSE and give more choices for path loss exponents.

This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

Hoang Thi Phuong Thao

Department of Electronics and Telecommunications, Electric Power University

No. 235 Hoang Quoc Viet Road, Hanoi, Vietnam

Email: thaohp@epu.edu.vn

1. INTRODUCTION

An increasing demand for communication traffic encourages the advent of fifth generation (5G) mobile technology in this period of global technological growth. As a result, technological innovations and additional frequency bands are proposed to accommodate the increased demand [1], [2]. As a result, 5G wireless communication is suggested and discussed. The 5G network is intended to meet three primary requirements: supporting machine communications applications in diverse industries, having a low reaction time to enable real-time communication, and effectively exploring available spectrum [3], [4]. The 5G technology outperforms in many ways, but the challenges posed by millimeter wave are critical. As a result, channel characterisation and measurement are observed, particularly wave attenuation studies at mm-wave frequency.

There have been many linear log-distance-based path loss models, such as those in [5]–[12], which are typically classified using one of two methods: empirical or deterministic [13]. Although the linear log-distance model is simple and tractable, the path loss prediction performance is not always accurate for all propagation environments. Therefore, it is necessary to propose methods for modeling and estimating pathloss with high performance for different contexts. In this paper, we create a wave propagation model for an urban region in

Hanoi, Vietnam, with the empirical model as our focus. Because empirical models are based on simulation and practical measurements in a specific scenario, and measured data may contain relationships or effects of propagation parameters with path loss values, it is a potential field for extracting, analyzing, and discovering the hidden structure of propagation model. Furthermore, due to its simplicity and lack of computational labor, a large number of empirical investigations and experiments have been launched to model wave propagation. The log-distance-based model is the most prevalent sort of propagating model. The NYU wireless campaign of real outdoor measurements developed two models to construct wave propagation, namely, the close-in (CI) reference distance model and the floating-intercept (FI) model for urban areas at frequencies of 28 GHz and 38 GHz [14], [15]. In addition, Patwari *et al.* [16] discusses and analyzes the log-distance based model. Machine learning is a data-driven research topic that enables machines or computers to learn without directly programming them.

Machine learning tasks are divided into three categories: supervised learning, unsupervised learning, and semi-supervised learning. Computers learn from a training set of labeled data and make judgments based on it. Unsupervised learning is also used to learn and find the unlabeled data's hidden and natural structure or pattern. Some papers have studied and presented applications of supervised machine learning techniques to deterministic wave propagation models [17]–[19]. However, it can occasionally overburden computing and make it impossible to apply to real-world scenarios. Besides, research into the data processing of empirical models is still in its early phases. The new Bayesian filtering method is used in the study proposed in [20] to estimate the pathloss exponent (PLE) of the log-normal shadowing propagation model for outdoor received signal strength indicator (RSSI) observations. The RSSI measurements were made using the Zigbee protocol and a broadcasting distance of 1 to 100 m. The method relies on weighted random samples (particles). Its execution mechanism is iterative and consists of three major stages: prediction, update, and resampling. In two innovative approaches, the publication [21] proposes the idea of noise floor for evaluating the path loss parameters of shortened data sets. In the case of an unknown number of samples below the noise floor, the PLE and standard deviation of large-scale fading are approximated using the likelihood equation for the truncated normal distribution. Meanwhile, if the number of missing samples is known, the estimation is based on an incomplete data set and the expectation maximization (EM) technique. The method was successfully used to data from automobiles communicating at 5.6 GHz. Aldossari and Chen [22] introduced a novel machine learning approach called multiple linear regression that can tackle the concerns of inaccuracy, complexity, and measurement capacity of traditional techniques. In comparison to conventional linear regression, multiple linear regression improves path loss model estimation.

In this research, we present a novel unsupervised machine learning processing model for the problem of path loss estimation in an empirical model. The proposed model is based on the cooperative and iterative evaluation exchange (Co-IEE) clustering algorithm's iterative method. Our model is capable of dealing with outliers in the data set. Furthermore, it can provide a plethora of scenarios or structures that may be buried in a particular data collection. To demonstrate the performance of the proposed model, data from wireless insite simulation of time city area-a typical urban area in Hanoi, Vietnam are analyzed, and the PLE and standard deviation are estimated using both the minimum mean square error (MMSE) approach and the suggested model.

2. CI FREE SPACE PATH LOSS MODEL

The CI free space reference distance path loss model [7], [23], [24] describes the dependence of path loss value on the physical distance between Tx-Rx. The CI model is given by (1):

$$PL(d)[dB] = FSPL + 10 \cdot \bar{n} \cdot \log_{10} \frac{d}{d_0} + X_{\sigma} \quad (1)$$

where FSPL denotes the free space path loss which can be calculated by Friis as (2):

$$FSPL(d)[dB] = 20 \cdot \log_{10} \frac{4\pi d_0 f}{c} \quad (2)$$

The environmental dependent parameter is the CI reference distance d_0 . This reference distance allows the propagating model to transmit the signal without attenuation or multipath fading. In other words, within this reference distance, the signal is transmitted in free space condition. In millimeter waveband, this reference distance d_0 is set to equal 1 m as standard because of its convenience in measurement and physical guarantee.

In CI model, since the FSPL just depends on frequency, the only parameter that need to be optimized is an overall path loss exponent \bar{n} . Since the PLE is indicated, the path loss value linearly depends on the quantity $\log_{10} \frac{d}{d_0}$. The PLE is a slope to measure the decreasing ratio of received signal strength (RSS) relied on the separating distance Tx-Rx. The increase in physical distance between Tx-Rx will lead to a lower RSS, or the higher value of path loss in propagating path. For the given data set of path loss value and corresponding Tx-Rx distance of receivers, we need to estimate the optimum PLE in order to the error (or loss function) between estimated set and the empirical one is minimum. X_σ is a Gaussian random variable with zero mean which is a quantity to represent the error (or the shadow fading) of function. The overall PLE is calculated based on linear regression, we need to find the best fitting line for the given data set. In this linear case, the variance of shadow fading is equal to the concept of error function named mean square error (MSE). The square root of MSE is root mean square error (RMSE) which is known as the standard deviation. The MSE and RMSE defined as (3):

$$\begin{aligned} MSE &= \frac{\sum(\hat{y} - y)^2}{2N} \\ RMSE &= \sqrt{MSE} \end{aligned} \quad (3)$$

Where \hat{y} is estimated value, y is sample one, and N is number of data points. Normally, the mean is halved as a convenience for the computation. In order to estimate the best fitting slope (or the best overall PLE), the classical regression algorithm usually used is gradient descent (GD). GD is the first-order iterative optimization algorithm of machine learning for finding the local minimum or global minimum of given linear function. The iterative process for converging to the minimum point of GD is given by (4):

$$x_{i+1} = x_i - \eta f'(x_i) \quad (4)$$

Where x_i is a point that we estimated after iteration i , while x_{i+1} is a point that we find after iteration $i + 1$; η is the learning rate which is given as input parameter of GD algorithm; and $f'(x_i)$ is derivative of considered function at x_i .

The GD algorithm starts at a random point, then executes its iterative procedures to gradually converge to the optimum point. To apply the GD algorithm to the problem of path loss estimation, we use the GD to iteratively minimize the MSE or RMSE of the data set having log-distance format as shown in (1). The overall PLE that allows the minimum MSE will be chosen as the result of estimation. In linear regression with one variable, the optimum point returned by GD is the global minimum, so the overall PLE returned by GD in this case will be the optimum one.

3. PROPOSED MODEL

Figure 1 illustrates the general flowchart of our proposed model for path loss estimation. It follows the procedure proposed in [25] with some additional steps. The proposed model includes six main steps: i) feature selection and conversion step extracts suitable and useful information from the raw data set, ii) forward evaluation exchange step calculates and transfers voting within receivers based on extracted data set, iii) backward evaluation exchange step calculates and transfers feedbacks, iv) convergence checking decides whether the procedure can stop or not; v) validating the clustering results, and vi) result interpretation and evaluation step chooses the best result to deduce the optimum PLE.

3.1. Data set

Our proposed model takes the raw data set of path loss model as its input, then executes its iterative procedure to finally deduce the overall PLE of given path loss model. Based on the proposed model, the simulated data which is researched in a specific area will be discovered and analyzed to reveal some useful patterns that can be used to deduce the optimum PLE of the considered path loss model. The raw data set includes some important features for estimating path loss model such as the physical distance between Tx-Rx, the path loss value of each receiver.

In CI model, the quantity $\log_{10} \frac{d}{d_0}$ is linearly affected to the value of path loss with the slope of overall PLE \bar{n} . It means that instead of finding the best fitting slope (overall PLE \bar{n}) with the raw data set as MMSE method did, our proposed model groups receivers that have similar PLE into clusters and then infer the most optimum overall PLE. The aim of our model is also to minimize the error between the estimated data with the raw one. For more detail, the act of grouping receivers leads to the aim of minimization inter-distance

between PLE of receivers. From there, we can evaluate the tendency of PLE values in considered data set, then deduce the overall PLE by considering the contributions of groups. Indicating an exact overall PLE based on tendencies of members' PLE allows the model to return the minimum error of estimated data.

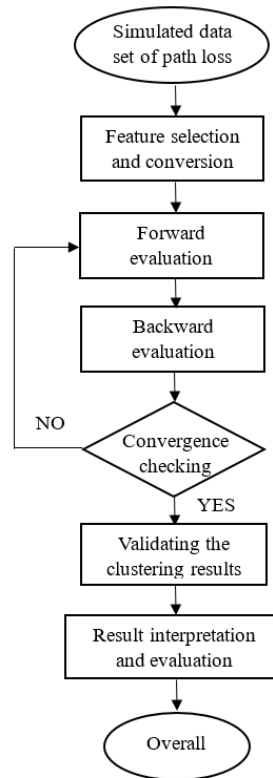


Figure 1. Flow chart of the proposed model

3.2. Feature selection and conversion

Overall, feature extraction step selects the core and general characteristics of objects based on criteria and rules. The selected characteristics would be converted and standardized into a general and consistent format. This step plays an important role in the model because it directly affects to the correctness and effectiveness of clustering applications in the next stage. In this study, our raw data set of wave attenuation is a set of received points, in which each point contains two data fields, i.e., physical separating distance Tx-Rx and path loss value which will be used to estimate the overall PLE. However, we need to transfer the raw data set into a suitable form, because the form of raw data would cause the wrong case when two receivers have the similar path loss exponents but the distance (or difference) between them is large because of the large difference between physical distances Tx-Rx of these two receivers and it will cause the large difference in received powers. In the CI path loss model, the goal of path loss estimation is to find the overall path loss exponent which allows the minimum error of estimated path loss compared to the practical one. To do this, our proposed model initially consider the specific path loss exponent of each received point which can be calculated from path loss value and distance Tx-Rx. Then the extracted data set will be analyzed and evaluated based on an iterative process which follows the procedure proposed by [25]. The iterative process includes 3 main phases, i.e., forward evaluation exchange, backward evaluation exchange, and convergence checking.

3.3. Forward evaluation exchange

Let consider a set of received points: $\{rx_1, rx_2, \dots, rx_N\}$. After the feature extraction step, each receiver has a general format as (5):

$$rx_i = [n_i] \in R, i = [1 : N] \quad (5)$$

Where n_i is a path loss exponent of received point rx_i . Then the model calculated the distance between each pair of receivers to prepare for forward evaluation calculation. Each distance is calculated based on the carrying information of receivers, the calculation is expressed as (6):

$$d_{i,k} = dk, i = |ni - nk|/2 \quad (6)$$

The distance between a pair of receivers reflects the level of their difference. A low distance implies that two receivers have similar path loss exponent and vice versa. In this phase, $N \times N$ forward evaluations are calculated and exchanged in parallel among N receivers. Each receiver rx_i will calculate and send N forward evaluations including $N - 1$ normal forward evaluations to other receivers and one self-forward evaluation to itself. Each normal forward evaluation sent from rx_i to rx_k ($i, k \in [1 : N]; i \neq k$) is calculated by (7):

$$f(i, k) = d(i, k) - m(i, k') \quad (7)$$

Where $m(i, k')$ is a “cumulative” distance between rx_i and its “closest” receiver except rx_k , in other words, $m(i, k')$ is a measurement of similarity between rx_i and the receiver (differs to rx_k) having most similar path loss exponent with rx_i . This quantity is expressed as (8):

$$m(i, k') = \min_{k' \neq k} \{d(i, k') + b_{pre}(k', i)\} \quad (8)$$

Where $d(i, k')$ is the distance between rx_i and $rx_{k'}$, $b_{pre}(k', i)$ is backward evaluation sent from $rx_{k'}$ to rx_i in the previous iteration, it is set to equal zero when the algorithm is running in the first iteration, and $b_{pre}(k', i)$ is the quantity representing the cumulative component in the calculation of $m(i, k')$.

The normal forward evaluation indicates the competitive ability of rx_i with other objects for being representative of rx_i . If $f(i, k) < 0$, the received point rx_k has ability to represent rx_i because there is no receiver being closer to rx_i than rx_k . In this case, the lower value of $f(i, k)$ indicates that rx_k has more ability to represent rx_i . In contrast, if $f(i, k) > 0$, the received point rx_k is not suitable to represent rx_i because it exists a receiver being closer to rx_i than rx_k . Besides, the self-forward evaluation that rx_i sent to itself is similarly calculated as the normal forward evaluation and shown in (7). It reflects the ability of rx_i for being representative based on the competition of rx_i with other receivers.

$$f(i, i) = p_i - \min_{k' \neq i} \{d(i, k') + b_{pre}(k', i)\} \quad (9)$$

p_i is an input quantity called preference value of each receiver rx_i . It directly affected to the competitive ability of receiver rx_i in the competition to become representative. In this study, because we aim to extract and discover the hidden structure of given data set and there is no existing information about the data set are known, so that we set preference values of all receivers are equal to a general preference value p . So, the self-forward evaluation of each receiver rx_i will be expressed as (10):

$$f(i, i) = p - \min_{k' \neq i} \{d(i, k') + b_{pre}(k', i)\} \quad (10)$$

The lower value of general preference p could cause the bigger number of clusters at the clustering output and vice versa. As a result, if the general preference value is low enough, it will lead to each object becomes a cluster, or the number of clusters is maximum. While a very high preference value causes the algorithm to return minimum number of clusters, or all receivers will live in one cluster. Based on this crucial characteristic, we can find the most accurate and suitable clustering results of given data set.

3.4. Backward evaluation exchange

As we have known in the first phase, the negative value of forward evaluation $f(i, k)$ reflects that rx_k is suitable for being representative of rx_i . Therefore, to evaluate the ability of being representative based on voting from other received points, the receiver rx_k calculates a quantity $b(k, k)$ which is a concept of self-backward evaluation and expressed as (11):

$$b(k, k) = \sum_{i' \neq k} \min\{0, f(i', k)\} \quad (11)$$

However, to fully evaluate its ability of being representative, rx_k adds the quantity $f(k, k)$ reflecting its competition with other objects to its supporting forward evaluations. The complete evaluation of rx_k is calculated by (12):

$$co(k, k) = f(k, k) + b(k, k) \quad (12)$$

In this phase, there are also $N \times N$ backward evaluations in total are calculated and exchanged in parallel among N objects. In particular, $N - 1$ normal backward evaluations are sent from each receiver to different ones and one self-backward is sent to itself. The concept of normal backward evaluation is calculated by (13):

$$b(k, i) = \max\{0, co(k, k) - \min\{0, f(i, k)\}\} \quad (13)$$

Because rx_k is sending backward evaluation to rx_i for reviewing, so the value of normal backward evaluation does not contain the voting from rx_i to rx_k . We set threshold to the value of normal backward evaluation $b(k, i)$ ($i \neq k$), so that it is not smaller than zero to avoid the situation when $b(k, i)$ is negative because of highly supported receiver rx_k which is not candidate representative of rx_i . This value of $b(k, i)$ not only leads to rx_i chooses rx_k as its representative while rx_k is not suitable to be the representative of rx_i , but also affects to the representative competition in the first phase of following iterations.

The positive value of $b(k, i)$ implies that rx_i is not suitable for being represented by rx_k because rx_k does not deserve as a representative (no one votes for rx_k), while the zero value of $b(k, i)$ implies that rx_i can be represented by rx_k because rx_k qualifies as a representative. It is important to note that the smaller value of self-backward evaluation reflects the higher ability of each rx_k to be representative, while the higher value of normal backward evaluation $b(k, i)$ reflects the lower ability of rx_i being represented by rx_k . While the exchange is completed, all backward evaluations will be updated as cumulative evidence for reviewing forward evaluations in the next iteration. There are $N \times N$ backward evaluations updated in total.

$$b_{pre}(k, i) = b(k, i) \text{ (where } i, k = [1 : N]) \quad (14)$$

3.5. Convergence checking

At the end of each iteration, the model always searches a representative for each receiver and check whether it is convergence or not. The representative c_i of each receiver rx_i is indicated based on (15):

$$c_i = \underset{o_k, k=[1:N]}{argmin} \{co(i, k)\} \quad (15)$$

Where $co(i, k)$ is the cooperative evaluation between rx_i and rx_k . It is a sum of forward evaluation sent from rx_i to rx_k and backward evaluation sent from rx_k back to rx_i . The cooperative evaluation $co(i, k)$ reflects the level of mutual suitability between rx_i and rx_k . The lower value of $co(i, k)$ reflects the higher level of suitability between rx_i and rx_k , and vice versa.

The model will follow that procedure until all representatives of N receivers are indicated. When the list of representatives of all considered receivers does not change after a given number of iterations, we enclose the iterative process and move to the step of validating results. The clustering output returned by the iterative process is the representative set of all N receivers. From the representative set returned by the model, we group receivers which have the same representative into a cluster and the representative of cluster is the general representative of all members in that cluster. Each receiver is located in only one cluster and the representative of each cluster is a receiver which has most similar carrying information to all members of that cluster, in other words it has the most similar path loss exponent with all members of that cluster. In this case study, there is no existing information about data set, the Co-IEE just works on its own way to discover the natural structure of the given data set. Therefore, Silhouette validity index, which is a kind of internal validation technique, is entirely suitable and applicable to assess the performance of clustering results. The silhouette index [26]-[28] of each receiver rx_i is defined as (16):

$$Sil(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (16)$$

Where $a(i)$ is the average distance between rx_i and other receivers locating in the same cluster with it, $b(i)$ is the minimum average distance between rx_i , and all other receivers in different cluster with rx_i . In other words,

$b(i)$ is the overall average distance between rx_i with its “closest” cluster. $a(i)$ and $b(i)$ are calculated based on euclidean distance to have synchronized unit with the input of Co-IEE.

To evaluate the overall performance of each clustering result, we use a concept of global silhouette which is defined as (17):

$$Sil(global) = \frac{\sum_{i=1}^N Sil_i}{N} \quad (17)$$

The single $Sil(i)$ reflects how well suited each receiver is when it locates in the current cluster. The value of $Sil(i)$ ranges from -1 to 1. The clustering result is good if the $Sil(i)$ is near 1 while the contrast validation occurs when $Sil(i)$ is near -1. The value of $Sil(i)$ is around 0 implies that rx_i hesitantly belongs to its current cluster and the nearest cluster with it. Meanwhile, the global silhouette Sil_{global} indicates the overall correctness and compactness of the clustering result. We can use this quantity to search for the most suitable number of clusters.

3.6. Result interpretation and evaluation

Based on the set of clusters formed in convergence checking step, we deduce an overall path loss exponent of the whole data set. To do this, we use the statistical concept of weighted arithmetic mean. The weighted mean indicates the center of a quantitative set of objects based on their different contributions in the set. Therefore, we can calculate the overall path loss exponent as shown in (18), where each quantitative object is a representative of each cluster, the weight is number of members in that cluster over N . Because the representative of each cluster is the most similar member to others in cluster, it can represent the contributions of its represented members.

$$\bar{n} = \frac{\sum_{i=1}^{noc} nom_i \cdot n_{c_i}}{N} \quad (18)$$

Where noc is number of clusters, nom_i is number of members in cluster i , and n_{c_i} is path loss exponent of the representative of cluster i .

4. SIMULATION RESULTS

4.1. Simulation scenario

To obtain a data set for path loss model, we chose time city—an urban area in Hanoi, Vietnam as a simulation location because this area is a typical place of urban environment with a lot of high-rise architecture. The simulation is implemented in wireless insite tool with the environmental setup approximating real life. The system is implemented with the transmitted power of 35 dBm and a communicating frequency of 28 GHz. The PLE and standard deviation are simulated and estimated by both the MMSE method and the proposed model.

4.2. Simulation results of minimum mean square error method

In order to compare the proposed model to the MMSE method, we first simulate PLE and RMSE using GD algorithm. Table 1 shows the estimated PLE and RMSE of simulated model of time city area. The implementations are executed in MATLAB tool with the learning rate of GD algorithm equal to 0.01 and the starting point of path loss exponent is 0.

Table 1. Estimated result of GD algorithm

	PLE (\bar{n})	RMSE (dB)
LOS	1.80	6.53
NLOS	3.34	35.94

From the estimated parameters shown in Table 1, we construct the system of equations for the CI model at 28 GHz as (19) and (20):

$$PL_{LOS}(d)[dB] = 61.385 + 18 \log \frac{d}{d_0} + X_{\sigma}(\text{with } \sigma = 6.53dB) \quad (19)$$

$$PL_{NLOS}(d)[dB] = 61.385 + 33.4 \log \frac{d}{d_0} + X_{\sigma}(\text{with } \sigma = 35.94dB) \quad (20)$$

Figures 2 and 3 show the MATLAB simulation results of path loss estimation using GD. In which each blue point is a receiver with corresponding path loss and the linear component $10 \log \frac{d}{d_0}$, while the red line is the best fitting line estimated by GD.

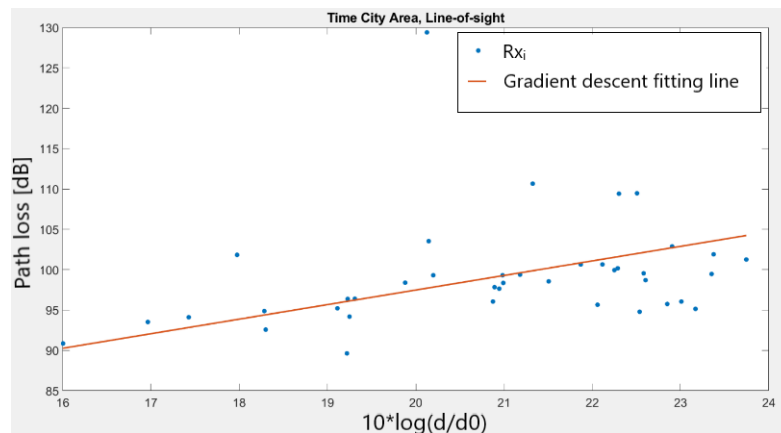


Figure 2. Path loss estimation of LOS data set

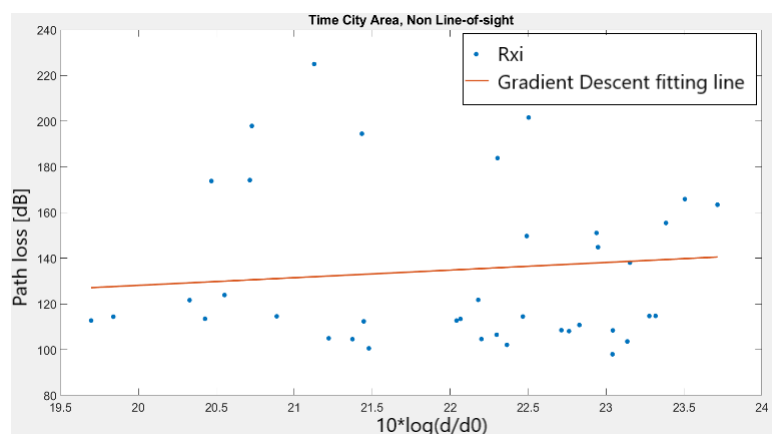


Figure 3. Path loss estimation of NLOS data set

4.3. Simulation results of proposed model

This subsection includes the simulated results of the proposed model applied to the data sets in time city area. There are two simulated data sets: the LOS set includes 40 receivers, and the NLOS set includes 40 receivers. The transmission is set at a frequency of 28 GHz, a transmitted power of 35 dBm, a transmitter's height of 7 m, and a receiver's height of 1.5 m. The simulation results of the relationship between the number of clusters, the global silhouette index, and the general preference value for the LOS data set and the NLOS data set are shown in following subsection.

4.3.1. Line-of-sight data set

Figures 4 and 5 illustrate the connections between general preference value with the number of clusters and global silhouette index respectively while the model runs on the LOS data set. As we can see in Figure 4, the model is most stable at 3 or 2 clusters, which corresponds to a range from 0.4642 to 0.9452 and a range from 0.9822 to 2.61. In Figure 5, we can see that the model returns the highest stable global silhouette index (0.9623) when the general preference value varies from 0.9822 to 2.61 (a range allows model to return 2 clusters). Besides, while the general preference value is smaller than 0.9822 (a range allows model to return 3 clusters), the values of global silhouette index are quite high but not very stable. Therefore, we focus on searching over the range from 0.9822 to 2.61 to find an optimum value of general preference value that allows model to return the lowest RMSE.

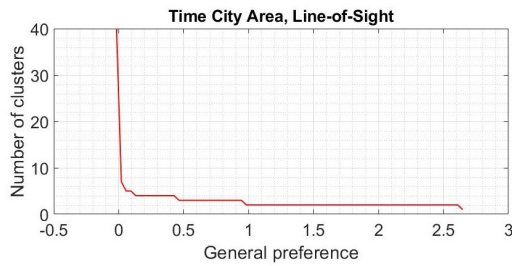


Figure 4. Graph of relationship between number of clusters and general preference value (LOS)

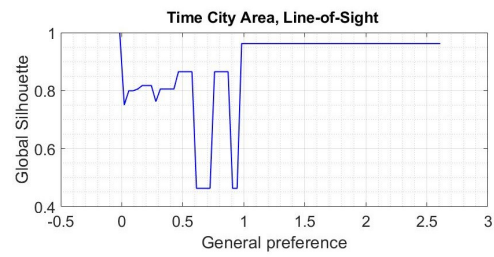


Figure 5. Graph of relationship between global silhouette and general preference value (LOS)

Figure 6 illustrates the connections of overall path loss exponent and corresponding RMSE with general preference value. It can be seen in Figure 6, the lowest value of RMSE that the model can reach is 6.526 dB. To reach this lowest RMSE value, the general preference value must be equal to a value belonging to a set $([0.9822, 1.093], 1.87, 2.055, 2.277, [2.425, 2.61])$. We can choose any value of this set for general preference value because all values of this set will allow model to return the same result. Moreover, as in the relationship of general preference with overall exponent, the estimated overall exponent equals to 1.801 when the result of model is optimum. Figure 7 provides the view of estimated result while the model is executed with specific general preference value of 0.9822. In which the star frame notes the cluster representatives, and the blue line is fitting line estimated by proposed model. We can see that the outliers (red points) are detected exactly by the model.

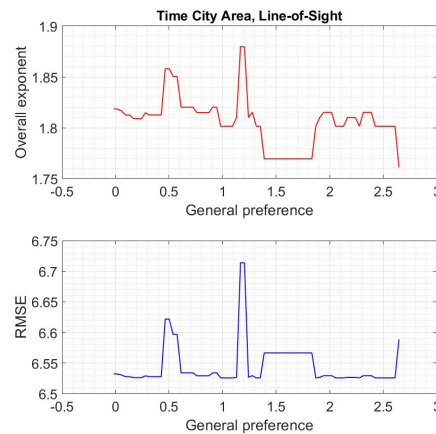


Figure 6. Connection between overall path loss exponent, RMSE and general preference value (LOS)

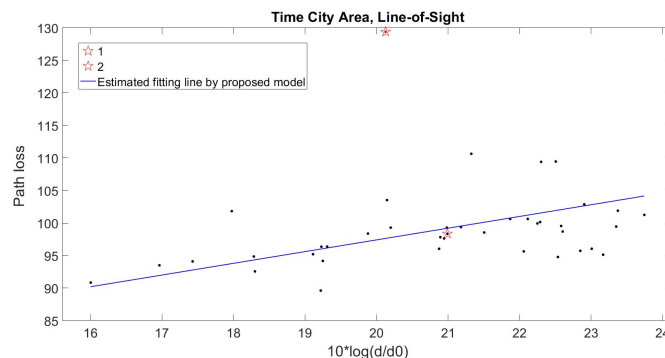


Figure 7. Visualization of estimated results returned with general preference of 0.9822 (LOS set)

4.3.2. Non line-of-sight data set

For the NLOS data set, the connections between general preference value with the number of clusters and global Silhouette index are shown in Figures 8 and 9. Overall, Figure 8 shows the view that the result of 2 clusters is the most stable scenario returned by the model, while the second one is the result of 3 clusters. These two results correspond to general preference values ranging from 16.9 to 86.62 and from 2.501 to 16.52, respectively. Figure 9 shows that when the general preference value varies from 2.501 to 86.62 (the range allows model to return 3 or 2 clusters), the global Silhouette values are quite high and vary from 0.81 to 0.86. Especially, the global silhouette value reaches the highest point overall, at 0.8604 while the preference values vary from 80.18 to 86.62. Therefore, we will search over the range from 2.501 to 86.62 of general preference value to choose the optimum value that allows model to return the best result.

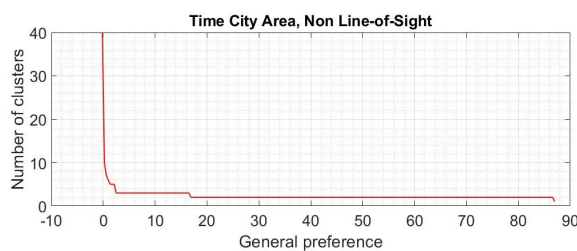


Figure 8. Graph of relationship between number of clusters and general preference value (NLOS)

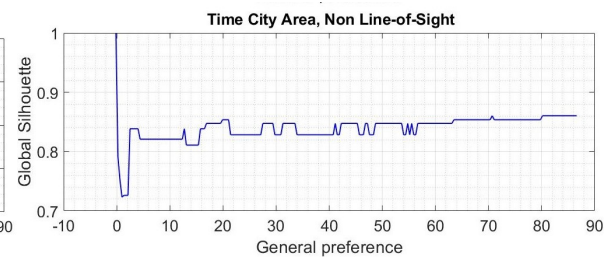


Figure 9. Graph of relationship between global silhouette and general preference value (NLOS)

Figure 10 provides a view of the connections between overall PLE and the corresponding RMSE with general preference value. As we can see in the graph of connection between general preference with RMSE, the lowest value of RMSE that the model can obtain is 35.94 dB. The result has this RMSE value when the general preference belongs to a set ([21.45, 27.13], [29.78, 30.92], [33.95, 63.12]). Then the value of overall path loss exponent could be equal to 3.344, 3.317, or 3.328. While the general preference value varies from 80.18 to 86.62, the RMSE is slightly higher than the lowest point, at 35.99 dB. Then the overall path loss exponent equals to 3.421. The estimated results by the works on model with general preference values of 21.45 and 80.18 are shown in Figures 11 and 12.

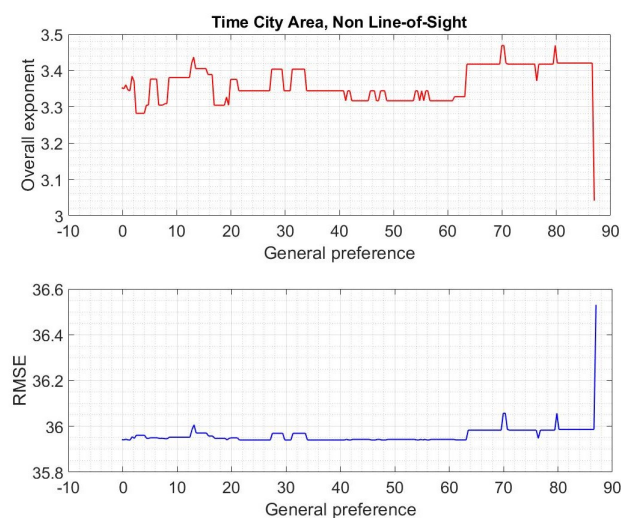


Figure 10. Connection between overall path loss exponent, RMSE, and general preference value (NLOS)

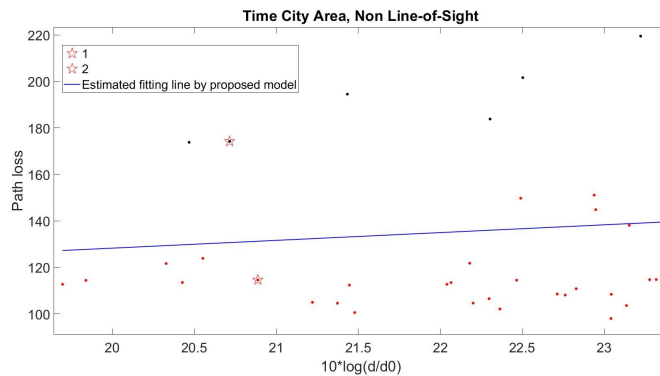


Figure 11. Visualization of estimated results with general preference of 21.45 (NLOS set)

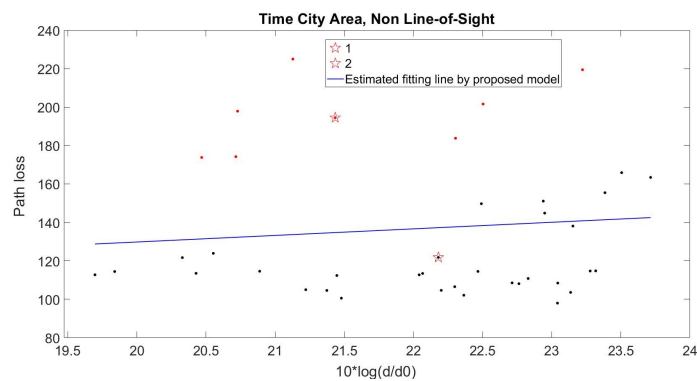


Figure 12. Visualization of estimated results with general preference of 80.18 (NLOS set)

4.4. Result comparison

Table 2 shows the estimated results of GD and the proposed model in time city area. As supported by Table 2, the path loss exponents of LOS condition estimated by two methods are equal with the same RMSE (or standard deviation). The differences between two methods are witnessed in the NLOS condition, in which the proposed model provides 4 optional path loss exponents with the same and slightly higher RMSE than the results of GD (35.94 dB and 35.99 dB). The only result of linear regression is an optimal one. It means that the result of path loss exponent estimated by GD is the best possible one. Therefore, the equivalent results of the proposed model with GD demonstrate that this model is correctly applied to the problem of path loss estimation with exact results.

Table 2. Result collections

	Data set	PLE(\bar{n})	RMSE (dB)
GD	LOS	1.80	6.53
	NLOS	3.34	35.94
Proposed model using Co-IEE	LOS	1.801	6.526
	NLOS	3.344	35.94
		3.317	35.94
		3.328	35.94
		3.421	35.99

5. CONCLUSION

In this study, we propose a novel model of signal processing that can be applied to millimeter wave propagation path loss model estimation. The simulation results prove that the model is successfully applied to the problem of path loss estimation. The path loss models of time city area (at 28 GHz, in both LOS and NLOS conditions) are then estimated by the proposed model. The results are compared to the optimum results of the

MMSE method and reflect the minor difference between them. For the LOS condition, the processing model returns the estimated PLE of 1.804 with the standard deviation of shadow fading equals to 6.526 dB. In NLOS condition, the PLE is estimated to equal 3.344, 3.317, or 3.328 with standard deviation of 35.94. Moreover, another possible scenario is returned by the proposed model with the PLE of 3.421 and standard deviation of 35.99 which has better correctness and separateness of clustering results. These results are similar to the theoretical path loss model, i.e., free space PLE of 2 and PLE in urban area ranging from 3 to 5. This signal processing model suggests the works of application to various path loss models and the in-depth discovery of the structure of path loss model, i.e., separating the data set for constructing a more detailed and fitting path loss model.




REFERENCES

- [1] T. S. Rappaport *et al.*, "Millimeter Wave Mobile Communications for 5G Cellular: It Will Work!," *IEEE Access*, vol. 1, pp. 335–349, 2013, doi: 10.1109/ACCESS.2013.2260813.
- [2] N. A. -Falahy, M. AlMahamy, and A. M. Mahmood, "Performance analysis of millimeter wave 5G networks for outdoor environment: propagation perspectives," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 20, no. 1, pp. 214–221, 2020, doi: 10.11591/ijeecs.v20.i1.pp214-221.
- [3] A. Osseiran *et al.*, "The Foundation of the Mobile and Wireless Communications System for 2020 and Beyond: Challenges, Enablers and Technology Solutions," in *2013 IEEE 77th Vehicular Technology Conference (VTC Spring)*, Dresden, Germany, 2013, pp. 1–5, doi: 10.1109/VTCSpring.2013.6692781.
- [4] M. M. Diallo, D. B. O. Konditi, and O. V. Bossou, "A miniaturized dual-band planar antenna with a square ring defected ground structure for 5G millimetre-wave applications," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 29, no. 1, pp. 197–205, 2022, doi: 10.11591/ijeecs.v29.i1.pp197-205.
- [5] M. Hata, "Empirical formula for propagation loss in land mobile radio services," *IEEE Transactions on Vehicular Technology*, vol. 29, no. 3, pp. 317–325, 1980, doi: 10.1109/T-VT.1980.23859.
- [6] S. Y. Seidel, T. S. Rappaport, S. Jain, M. L. Lord, and R. Singh, "Path loss, scattering and multipath delay statistics in four European cities for digital cellular and microcellular radiotelephone," *IEEE Transactions on Vehicular Technology*, vol. 40, no. 4, pp. 721–730, 1991, doi: 10.1109/25.108383.
- [7] Y. Okumura, E. Ohmori, T. Kawano, and K. Fukuda, "Field strength and its variability in VHF and UHF land-mobile radio service," *Review of the Electrical Communication Laboratory*, vol. 16, no. 9–10, pp. 825–873, 1968.
- [8] P. E. Mogensen and J. Wigard, COST Action 231: Digital Mobile Radio Towards Future Generation Systems: Final Report, vol. 18957. Bruxelles, Belgium: European Commission, 1999.
- [9] A. Akeyama, T. Nagatsu, and Y. Ebine, "Mobile radio propagation characteristics and radio zone design method in local cities," *Review of the Electrical Communication Laboratory*, vol. 30, pp. 309–317, 1982.
- [10] W. Tang, X. Ma, J. Wei, and Z. Wang, "Measurement and Analysis of Near-Ground Propagation Models under Different Terrains for Wireless Sensor Networks," *Sensors*, vol. 19, no. 8, Jan. 2019, doi: 10.3390/s19081901.
- [11] J. Bi *et al.*, "Fast Radio Map Construction by using Adaptive Path Loss Model Interpolation in Large-Scale Building," *Sensors*, vol. 19, no. 3, pp. 1–19, 2019, doi: 10.3390/s19030712.
- [12] A. T. A. -Heety, M. T. Islam, A. H. Rashid, H. N. A. Ali, A. M. Fadil, and F. Arabian, "Performance evaluation of wireless data traffic in mm Wave massive MIMO communication," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 20, no. 3, pp. 1342–1350, 2020, doi: 10.11591/ijeecs.v20.i3.pp1342-1350.
- [13] E. Ostlin, H. -J. Zepernick, and H. Suzuki, "Macrocell radio wave propagation prediction using an artificial neural network," in *IEEE 60th Vehicular Technology Conference, 2004. VTC2004-Fall. 2004, Los Angeles, CA, USA, 2004*, pp. 57–61, doi: 10.1109/VETECF.2004.1399921.
- [14] G. R. MacCartney, J. Zhang, S. Nie, and T. S. Rappaport, "Path loss models for 5G millimeter wave propagation channels in urban microcells," in *2013 IEEE Global Communications Conference (GLOBECOM)*, Atlanta, GA, 2013, pp. 3948–3953, doi: 10.1109/GLOCOM.2013.6831690.
- [15] Y. Azar *et al.*, "28 GHz propagation measurements for outdoor cellular communications using steerable beam antennas in New York city," in *2013 IEEE International Conference on Communications (ICC)*, Budapest, Hungary, 2013, pp. 5143–5147, doi: 10.1109/ICC.2013.6655399.
- [16] N. Patwari, A. O. Hero, M. Perkins, N. S. Correal, and R. J. O'Dea, "Relative location estimation in wireless sensor networks," *IEEE Transactions on Signal Processing*, vol. 51, no. 8, pp. 2137–2148, 2003, doi: 10.1109/TSP.2003.814469.
- [17] H. -S. Jo, C. Park, E. Lee, H. K. Choi, and J. Park, "Path Loss Prediction Based on Machine Learning Techniques: Principal Component Analysis, Artificial Neural Network, and Gaussian Process," *Sensors*, vol. 20, no. 7, pp. 1–23, 2020, doi: 10.3390/s20071927.
- [18] Y. Zhang, J. Wen, G. Yang, Z. He, and J. Wang, "Path Loss Prediction Based on Machine Learning: Principle, Method, and Data Expansion," *Applied Sciences*, vol. 9, no. 9, pp. 1–18, 2019, doi: 10.3390/app9091908.
- [19] E. Ostlin, H. -J. Zepernick, and H. Suzuki, "Macrocell Path-Loss Prediction Using Artificial Neural Networks," *IEEE Transactions on Vehicular Technology*, vol. 59, no. 6, pp. 2735–2747, 2010, doi: 10.1109/TVT.2010.2050502.
- [20] P. Wojcik, T. Zientarski, M. Charytanowicz, and E. Lukasik, "Estimation of the Path-Loss Exponent by Bayesian Filtering Method," *Sensors*, vol. 21, no. 6, pp. 1–11, 2021, doi: 10.3390/s21061934.
- [21] A. Taimoor, C. Gustafson, and F. Tufvesson, "Pathloss Estimation Techniques for Incomplete Channel Measurement Data," *COST IC1004 10th Management Committee and Scientific Meeting*, pp. 1–7, 2014.
- [22] S. Aldossari and K. -C. Chen, "Predicting the Path Loss of Wireless Channel Models Using Machine Learning Techniques in MmWave Urban Communications," in *2019 22nd International Symposium on Wireless Personal Multimedia Communications (WPMC)*, Lisbon, Portugal, 2019, pp. 1–6, doi: 10.1109/WPMC48795.2019.9096057.




- [23] S. Sun *et al.*, “Propagation Path Loss Models for 5G Urban Micro- and Macro-Cellular Scenarios,” in *2016 IEEE 83rd Vehicular Technology Conference (VTC Spring)*, Nanjing, China, 2016, pp. 1–6, doi: 10.1109/VTCSpring.2016.7504435.
- [24] S. Piersanti, L. A. Annoni, and D. Cassioli, “Millimeter waves channel measurements and path loss models,” in *2012 IEEE International Conference on Communications (ICC)*, Ottawa, ON, Canada, 2012, pp. 4552–4556, doi: 10.1109/ICC.2012.6363950.
- [25] D. H. Duc, C. Dang, T. T. Thao, V. V. Yem, and H. M. Son, “A Novel Clustering Algorithm based on Cooperative and Iterative Evaluation Exchange for University Classification,” in *2020 IEEE Eighth International Conference on Communications and Electronics (ICCE)*, Phu Quoc Island, Vietnam, 2021, pp. 307–312, doi: 10.1109/ICCE48956.2021.9352112.
- [26] P. J. Rousseeuw, “Silhouettes: A graphical aid to the interpretation and validation of cluster analysis,” *Journal of Computational and Applied Mathematics*, vol. 20, pp. 53–65, 1987, doi: 10.1016/0377-0427(87)90125-7.
- [27] A. F. Moiane and Á. M. L. Machado, “Evaluation of the clustering performance of affinity propagation algorithm considering the influence of preference parameter and damping factor,” *Boletim de Ciências Geodésicas*, vol. 24, no. 4, pp. 426–441, 2018, doi: 10.1590/s1982-21702018000400027.
- [28] T. Thinsungnoen, N. Kaoungku, P. Durongdumronchai, K. Kerdprasop, and N. Kerdprasop, “The Clustering Validity with Silhouette and Sum of Squared Errors,” in *The 3rd International Conference on Industrial Application Engineering 2015 (ICIAE2015)*, 2015, pp. 44–51, doi: 10.12792/iciae2015.012.

BIOGRAPHIES OF AUTHORS






Vu Thanh Quang    received the bachelor's degree and the master's degree in Electronics and Telecommunications in 1997. Currently, he is working at the voice of Vietnam. His research interests are radio and television systems, positioning systems, and next-generation wireless networks. He can be contacted at email: vuparis75013@gmail.com.






Do Huu Duc    received bachelor of arts (Second Class Hons.) was born in Vietnam, in 1998. He received the diploma of engineer in Electronics and Telecommunications from Hanoi University of Science and Technology, Vietnam in 2021. Currently, he is a master student at École Normale Supérieure Paris-Saclay. His research interests are data mining and signal processing in advanced wireless communication systems. He can be contacted at email: dhd131198@gmail.com.



Vu Van Yem    received bachelor of arts (Second Class Hons.) was born in Vietnam in 1975. He received the bachelor's and master's degree from the Hanoi University of Science and Technology. He was awarded doctor of philosophy in Electronics and Telecommunication Engineering from Télécom Paris Tech in 2005. He is currently working as a full professor in the School of Electronics and Telecommunications, Hanoi University of Science and Technology. His main research interests are multi-antenna wireless communication systems, antennas, and ultra-high frequency techniques. He can be contacted at email: yem.vuvan@hust.edu.vn.



Hoang Thi Phuong Thao    was born in Vietnam, in 1981. She received the diploma of engineer (2004), master of science (2007), and Ph.D. degree (2019) in Electronics and Telecommunications from Hanoi University of Science and Technology, Vietnam. Currently, she is a lecturer at Faculty of Electronics and Telecommunications, Electric Power University, Vietnam. Her research interests include designing antennas, localization systems, and wireless communication systems. She can be contacted at email: thaohp@epu.edu.vn.